Security analysis of control systems under nonlinear attacks
(extended abstract)

Ruslan Seifullaev[1] and André M. H. Teixeira[1]

***Index Terms*— Secure control, nonlinear systems, absolute stability, dissipativity, output-to-output gain**

Networked control systems (NCS) have recently attracted considerable attention for their capability to integrate distributed sensors, actuators, and controllers over a shared network infrastructure. This integration enables distributed control, allowing system components to be physically separated while still functioning as a cohesive unit. NCSs provide multiple advantages, including reduced wiring complexity, enhanced flexibility, and scalability. Despite these benefits, incorporating wireless communication networks into control loops introduces several challenges and constraints, such as sampled data, narrow bandwidth, latency, fading, interference, packet dropouts, etc. In addition, in recent years, the growing threat of cyberattacks has made NCS security a major concern [1], [2], [3]. Malicious activities such as malware injection and encryption key theft can compromise data integrity, grant unauthorized access to remote control centers, and ultimately degrade control loop performance or even lead to instability and failure. Addressing these threats requires a deep understanding of attack processes and their consequences, as well as the adaptation of network control strategies to mitigate them.

In this abstract, we consider nonlinear attacks, in which an adversary accesses sensor data and applies a nonlinear transformation to the measurement output. Consequently, the closed-loop system becomes nonlinear, making stability analysis nontrivial. Assuming that the nonlinearity belongs to a class of local or integral quadratic constraints, we provide frequency-domain conditions that guarantee absolute stability, based on the Yakubovich quadratic criterion [4], [5]. Additionally, to minimize the impact of stealthy attacks, where an adversary aims to maximize their effect on system performance by injecting an additive signal while remaining undetected, we analyze the output-to-output gain [6], a security metric that combines both performance impact and attack detectability. Using dissipativity theory, we derive a computationally efficient approach for estimating this metric.

We consider the following system:

$$\begin{aligned} \dot{x}_{\mathrm{p}}(t) &= A_{\mathrm{p}}x_{\mathrm{p}}(t) + B_{\mathrm{p}}u(t), \\ y_{\mathrm{m}}(t) &= C_{\mathrm{m,o}}x_{\mathrm{p}}(t), \;\; y_{\mathrm{p}}(t) = C_{\mathrm{p,o}}x_{\mathrm{p}}(t) + D_{\mathrm{p,o}}u(t), \end{aligned} \tag{1}$$

where $x_{\mathrm{p}}(t) \in \mathbb{R}^{n_x}$ is the state vector, $u(t) \in \mathbb{R}^{n_u}$ is the control input, $y_{\mathrm{m}}(t) \in \mathbb{R}^{n_y}$ is the measurement output, $y_{\mathrm{p}} \in \mathbb{R}^{n_{y_{\mathrm{p}}}}$ is the performance output, and A_{p}, B_{p}, $C_{\mathrm{m,o}}$, $C_{\mathrm{p,o}}$ and $D_{\mathrm{p,o}}$ are the matrices of appropriate dimensions, the pair $(A_{\mathrm{p}}, B_{\mathrm{p}})$ is controllable and the pair $(A_{\mathrm{p}}, C_{\mathrm{m,o}})$ is observable. Assume that the system is under a malicious attack, where an adversary can access the sensor measurements and replace the vector $y_{\mathrm{m}}(t)$ with $\tilde{y}_{\mathrm{m}}(t)$.

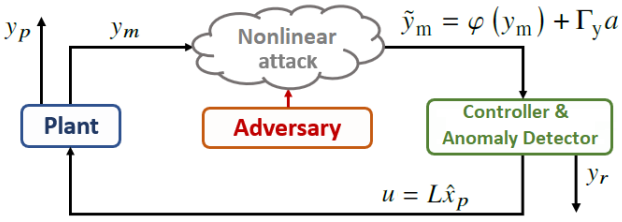

Fig. 1: Nonlinear attack

To estimate the state vector $x_{\mathrm{p}}(t)$ and design a feedback based on it, we use the following observer-based state feedback controller:

$$\begin{aligned} \dot{\hat{x}}_{\mathrm{p}}(t) &= A_{\mathrm{p}}\hat{x}_{\mathrm{p}}(t) + B_{\mathrm{p}}u(t) + Ky_{\mathrm{r}}(t), \quad u(t) = L\hat{x}_{\mathrm{p}}(t), \\ \hat{y}_{\mathrm{m}}(t) &= C_{\mathrm{m,o}}\hat{x}_{\mathrm{p}}(t), \quad y_{\mathrm{r}}(t) = \tilde{y}_{\mathrm{m}}(t) - \hat{y}_{\mathrm{m}}(t), \end{aligned} \tag{2}$$

where $K \in \mathbb{R}^{n_x \times n_y}$ and $L \in \mathbb{R}^{n_u \times n_x}$ are the observer and controller gains, respectively. Note that we use its residual output $y_{\mathrm{r}}(t)$, the difference between the measured and predicted outputs, for an anomaly detector, which raises an alarm if the residual becomes abnormally high. The adversary's goal is to degrade the performance, measured by $||y_{\mathrm{p}}||_{L_2}$, while simultaneously remaining undetected, i.e., keeping $||y_{\mathrm{r}}||_{L_2}$ small. The adversary can achieve this by applying the nonlinear transformation φ to the measured output y_{m} while also injecting an additive signal $\Delta y = \Gamma_{\mathrm{y}} a(t)$ to the sensors' output:

$$\tilde{y}_{\mathrm{m}}(t) = \varphi\left(y_{\mathrm{m}}(t), t\right) + \Gamma_{\mathrm{y}} a(t), \tag{3}$$

where $a(t) \in \mathbb{R}^{n_a}$, $\Gamma_{\mathrm{y}} \in \mathbb{R}^{n_y \times n_a}$, see Fig. 1. Then, the closed-loop system (1) – (3) takes the following form:

$$\begin{aligned} \dot{x}(t) &= Ax(t) + q\xi(t) + Ba(t), \\ \xi(t) &= \varphi(\sigma(t), t), \quad \sigma(t) = C_{\mathrm{m}}x(t) \\ y_{\mathrm{p}}(t) &= C_{\mathrm{p}}x(t), \quad y_{\mathrm{r}}(t) = C_{\mathrm{r}}x(t) + \xi(t) + D_{\mathrm{r}}a(t), \end{aligned} \tag{4}$$

where $x = \begin{bmatrix} x_{\mathrm{p}} \\ x_{\mathrm{p}} - \hat{x}_{\mathrm{p}} \end{bmatrix}$, and $A = \begin{bmatrix} A_{\mathrm{p}} + B_{\mathrm{p}}L & -B_{\mathrm{p}}L \\ KC_{\mathrm{m,o}} & A_{\mathrm{p}} - KC_{\mathrm{m,o}} \end{bmatrix}$, $q = \begin{bmatrix} 0 \\ -K \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ -K\Gamma_{\mathrm{y}} \end{bmatrix}$, $C_{\mathrm{m}} = \begin{bmatrix} C_{\mathrm{m,o}}, & 0 \end{bmatrix}$, $C_{\mathrm{p}} = \begin{bmatrix} C_{\mathrm{p,o}} + D_{\mathrm{p,o}}L, & -D_{\mathrm{p,o}}L \end{bmatrix}$, $C_{\mathrm{r}} = \begin{bmatrix} -C_{\mathrm{m,o}}, & C_{\mathrm{m,o}} \end{bmatrix}$, $D_{\mathrm{r}} = \Gamma_{\mathrm{y}}$.

[1]Ruslan Seifullaev is a postdoctoral researcher, and André M. H. Teixeira is an associate professor in the Division of Systems and Control, Department of Information Technology, Uppsala University, Sweden. Email addresses: {ruslan.seifullaev, andre.teixeira}@it.uu.se

A. Absolute stability

In this section, we provide frequency-domain conditions guaranteeing absolute stability of the closed-loop system (4) with $a(t) \equiv 0$. We assume that the nonlinearity φ belongs to some class $\mathfrak{M}_F = \{(\sigma(t), \xi(t))\}$, such that the functions $\xi(t)$ and $\sigma(t)$ satisfy the quadratic constraints

$$F\left(\xi(t), \sigma(t)\right) \geq 0$$

for all $t \geq 0$, where F is a quadratic form with a symmetric matrix $\bar{F} = \begin{bmatrix} F_{11} & F_{12} \\ * & F_{22} \end{bmatrix} \in \mathbb{R}^{2n_y \times 2n_y}$. Note that the results below are also valid for integral quadratic constraints (IQCs), where the class $\mathfrak{M}_F$ is such that $\int_0^\infty F\left(\xi(t), \sigma(t)\right) dt \geq 0$.

Definition 1: The closed-loop system (4) is called *minimally stable* in a class $\mathfrak{M}_F$ if there exist a solution $x(t)$ satisfying $(\sigma(t), \xi(t)) \in \mathfrak{M}_F$ such that $\lim_{t\to\infty} \|x(t)\| = 0$.

Definition 2: The closed-loop system (4) is called *absolutely stable* in a class $\mathfrak{M}_F$ if for any solution $x(t)$ satisfying $(\sigma(t), \xi(t)) \in \mathfrak{M}_F$ there exists a constant $c_1 > 0$ such that $\|x\|_{L_2}^2 + \|\xi\|_{L_2}^2 \leq c_1 \|x(0)\|^2$.

Theorem 1: Assume that the matrix $C_{\mathrm{m}}(sI - A)^{-1}q$ does not have poles on the imaginary axis, and the closed-loop system (4) is minimally stable in the class $\mathfrak{M}_F$. Then it is absolutely stable if

$$\tilde{F}(i\omega, \tilde{\xi}) < 0, \quad \text{for all } \omega \in \mathbb{R} \text{ and } \tilde{\xi} \in \mathbb{C}^{n_y}, \tilde{\xi} \neq 0, \tag{5}$$

where the Hermitian form $\tilde{F}$ is the extension of the quadratic form F obtained as $\tilde{F}(s, \tilde{\xi}) = F(\tilde{\xi}, -G_{\xi\sigma}(s)\tilde{\xi})$.

B. Output-to-output gain

A security metric that combines both performance impact and attack detectability was introduced in [6], [2], termed the output-to-output gain (OOG). The OOG metric is formulated as the optimal control problem:

$$OOG \triangleq \sup_{a \in L_{2e}, \varphi \in \mathfrak{M}_F} ||y_{\mathrm{p}}||_{L_2}^2, \quad \text{s.t.} \quad ||y_{\mathrm{r}}||_{L_2}^2 \leq 1,\ x(0) = 0,$$

where $L_{2e} = \left\{ a : \mathbb{R}_+ \to \mathbb{R}^{n_a} \,\middle|\, ||a||_{L_{2[0,T]}} < \infty,\ \forall T < \infty \right\}$. In other words, the OOG characterizes the adversary's goal of achieving maximum impact while avoiding detection. The following theorem provides a computational approach to estimating the OOG. The resulting linear matrix inequalities (LMIs) ensure that the system (4) is strictly dissipative [7], which, in turn, guarantees the boundedness of the OOG.

Theorem 2: Assume that there exist a matrix $P \geq 0$ and scalars $\kappa \geq 0$ and $\gamma > 0$ such that the LMI

$$\Psi_0(P) + \Psi_1(\gamma) + \Psi_2(\kappa) \leq 0 \tag{6}$$

is feasible, where $\Psi_1(\gamma) = \begin{bmatrix} \gamma C_{\mathrm{p}}^{\mathrm{T}} C_{\mathrm{p}} - C_{\mathrm{r}}^{\mathrm{T}} C_{\mathrm{r}} & -C_{\mathrm{r}}^{\mathrm{T}} q_{\mathrm{r}} & -C_{\mathrm{r}}^{\mathrm{T}} D_{\mathrm{r}} \\ * & -q_{\mathrm{r}}^{\mathrm{T}} q_{\mathrm{r}} & -q_{\mathrm{r}}^{\mathrm{T}} D_{\mathrm{r}} \\ * & * & -D_{\mathrm{r}}^{\mathrm{T}} D_{\mathrm{r}} \end{bmatrix}$,

$\Psi_0(P) = \begin{bmatrix} A^{\mathrm{T}}P + PA & Pq & PB \\ * & 0 & 0 \\ * & * & 0 \end{bmatrix}$, and $\Psi_2(\kappa) = \kappa \begin{bmatrix} r^{\mathrm{T}} F_{22} r & r^{\mathrm{T}} F_{12}^{\mathrm{T}} & 0 \\ * & F_{11} & 0 \\ * & * & 0 \end{bmatrix}$.

Then $OOG \leq \frac{1}{\gamma}$.

Finally, using the Kalman-Yakubovich-Popov (KYP) lemma [8], [9], we provide frequency domain conditions that guarantee the upper bound for the OOG. Introduce the following transfer matrices:

$$\begin{aligned}
G_{a\mathrm{p}}(s) &= C_{\mathrm{p}}(sI - A)^{-1}B, \quad G_{a\mathrm{r}}(s) = C_{\mathrm{r}}(sI - A)^{-1}B + D_r, \\
G_{\xi\mathrm{p}}(s) &= C_{\mathrm{p}}(sI - A)^{-1}q, \quad G_{\xi\mathrm{r}}(s) = C_{\mathrm{r}}(sI - A)^{-1}q + I, \\
G_{a\sigma}(s) &= C_{\mathrm{m}}(sI - A)^{-1}B, \quad G_{\xi\sigma}(s) = C_{\mathrm{m}}(sI - A)^{-1}q.
\end{aligned}$$

Define also the matrix

$$\Psi(s, \gamma, \kappa) = \begin{bmatrix} \Psi_{11}(s, \gamma, \kappa) & \Psi_{12}(s, \gamma, \kappa) \\ * & \Psi_{22}(s, \gamma, \kappa) \end{bmatrix},$$

where

$$\begin{aligned}
\Psi_{11}(s, \gamma, \kappa) &= \gamma G_{\xi\mathrm{p}}^*(s) G_{\xi\mathrm{p}}(s) - G_{\xi\mathrm{r}}^*(s) G_{\xi\mathrm{r}}(s) \\
&\quad + \kappa F_{12} G_{\xi\sigma}(s) + \kappa G_{\xi\sigma}^*(s) F_{12}^{\mathrm{T}} \\
&\quad + \kappa G_{\xi\sigma}^*(s) F_{22} G_{\xi\sigma}(s) + \kappa F_{11}, \\
\Psi_{12}(s, \gamma, \kappa) &= \gamma G_{\xi\mathrm{p}}^*(s) G_{a\mathrm{p}}(s) - G_{\xi\mathrm{r}}^*(s) G_{a\mathrm{r}}(s) \\
&\quad + \kappa G_{\xi\sigma}^*(s) F_{22} G_{a\sigma}(s) + \kappa F_{12} G_{a\sigma}(s), \\
\Psi_{22}(s, \gamma, \kappa) &= \gamma G_{a\mathrm{p}}^*(s) G_{a\mathrm{p}}(s) - G_{a\mathrm{r}}^*(s) G_{a\mathrm{r}}(s) \\
&\quad + \kappa G_{a\sigma}^*(s) F_{22} G_{a\sigma}(s).
\end{aligned}$$

Theorem 3: Assume that the matrix A is Hurwitz stable, the pair $(A, [q\ B])$ is controllable, and there exist scalars $\kappa \geq 0$ and $\gamma > 0$ such that

$$\Psi(i\omega, \gamma, \kappa) \leq 0 \quad \forall \omega \in \mathbb{R}, \tag{7}$$

and

$$\gamma C_{\mathrm{p}}^{\mathrm{T}} C_{\mathrm{p}} - C_{\mathrm{r}}^{\mathrm{T}} C_{\mathrm{r}} + \kappa r^{\mathrm{T}} F_{22} r \geq 0. \tag{8}$$

Then $OOG \leq \frac{1}{\gamma}$.

Remark 1: If the pair $(A, [q\ B])$ is controllable and $\det(i\omega I - A) \neq 0$, then, from the KYP lemma, it follows that the condition (7) is necessary for the feasibility of the LMI (6).

References

[1] A. A. Cardenas, S. Amin, and S. Sastry, "Secure control: Towards survivable cyber-physical systems," in *the 28th International Conference on Distributed Computing Systems Workshops*, 2008, pp. 495–500.

[2] A. M. H. Teixeira, "Security metrics for control systems," in *Safety, Security and Privacy for Cyber-Physical Systems*, R. M. Ferrari and A. M. H. Teixeira, Eds. Springer International Publishing, 2021.

[3] R. Seifullaev, A. M. H. Teixeira, and A. Ahlén, "Event-triggered control of nonlinear systems under deception and denial-of-service attacks," in *63rd IEEE Conference on Decision and Control (CDC)*, 2024, pp. 940–947.

[4] V. Yakubovich, "A quadratic criterion for absolute stability," in *Doklady Mathematics*, vol. 58, no. 1, 1998, pp. 169–172.

[5] A. Proskurnikov, "The Yakubovich quadratic criterion, F-stability and multi-agent consensus," *IFAC-PapersOnLine*, vol. 48, no. 11, pp. 414–419, 2015.

[6] A. M. H. Teixeira, H. Sandberg, and K. H. Johansson, "Strategic stealthy attacks: The output-to-output l_2-gain," in *54th IEEE Conference on Decision and Control (CDC)*, 2015, pp. 2582–2587.

[7] H. K. Khalil, *Nonlinear Systems*. Prentice Hall PTR, 2002.

[8] V. A. Yakubovich, "Solution of certain matrix inequalities in the stability theory of nonlinear control systems (english translation)," *Soviet Math. Dokl.*, vol. 3, 1962.

[9] A. Rantzer, "On the Kalman–Yakubovich–Popov lemma," *Systems & Control Letters*, vol. 28, no. 1, pp. 7–10, 1996.